A 20-Year Perspective on Incentive Regulation for Public Utilities

Regulation and Investment Conference
Australian Competition and Consumer Commission

Sydney, March 26/27, 2001

By Ingo Vogelsang[*)]

Abstract

Over the last 20 years, incentives have breathed new life into public utility regulation. Price caps have been the main force for end-user prices and benchmark regulation for essential inputs. The beauty of price caps is that they successfully combine incentives for cost reduction with incentives to rebalance prices towards allocative efficiency. The flexibility and incentives provided also make them an excellent tool when public utility sectors are opened to competition. We argue that price caps have properties leading to the right amount of competition, when it is unknown if the underlying market structure is naturally oligopolistic or monopolistic. This is of particular importance, because empirical work on natural market structures is largely inconclusive. While price caps make a regulated incumbent more aggressive in potentially competitive markets, this aggression is likely to improve on the unregulated outcome. There is, however, a proviso on the possibility of anticompetitive behavior. This is probably one reason why prices for essential inputs are often regulated on the basis of benchmarked costs. Such regulation can increase downstream competition and provide good incentives for minimizing total industry costs. The allocative efficiency properties depend on the amount and type of competition. As the market share of rivals increases, a case can be made for an increase of essential input prices relative to incremental costs. Benchmarked costs should, over time, lead into input price caps. While it could increase allocative efficiency if those were combined with output price caps to form global price caps, we believe that partial deregulation of end-user prices is to be preferred.

[*)] Boston University. The author would like to thank Bridger Mitchell for helpful comments.

Introduction: The Roots of Incentive Regulation

By the end of the 1970s economists had no doubt that regulation of competitive industries was inappropriate and a major failure. The successful deregulation of airlines and trucking proves the case, although some regulation of competitive industries, for example of taxicabs, survived. The regulation of public utilities was viewed differently, although the leading practice of rate-of-return regulation had been severely criticized at least since the early 1960s, when the discovery of the Averch-Johnson effect and the empirical work of Stigler and Friedland had shown the lack of improvement of regulation over unregulated monopoly outcomes.[1] It is thus not surprising that the same time that brought us deregulation also set the stage for new developments of public utility regulation. The critique of traditional regulation naturally induced economists both to recommend deregulation and to look for improvements of regulation. What is now known as incentive regulation has breathed new life into the stale public utility regulation. What do we mean by incentive regulation? In particular, it means that the regulator delegates certain performance-related decisions to the firm and that the profits of the regulated firm depend on performance measures of the regulator. Incentive regulation makes use of the firm's information advantage. The regulator thus controls less behavior but rather rewards outcomes.

Worldwide, the introduction of incentive regulation has been part of the regulatory reform movement, consisting of privatization, liberalization and deregulation. Incentive regulation is less radical than the other three. It can be viewed as a move towards deregulation, because it frees regulated firms from some rigid constraints. Being less radical its political and organizational implementation may not require the same kind of political and market changes as the other three. However, in practice, it has largely been linked to them. In the U.K., incentive regulation has accompanied privatization and liberalization in the main public utility industries, such as telecommunications, gas, water, and electricity. In the U.S., incentive regulation has facilitated liberalization and partial deregulation in telecommunications. Because of preexisting rate-of-return regulation the shift to incentive regulation was harder in the U.S. and led to more compromises than in countries with less regulatory traditions. Also, U.S. incentive regulation is more diverse than in other countries (Crew and Kleindorfer, 1996). Incentive regulation flourishes in the same political climate and ideology as those more radical types of sector reform. This has both to do with the delegation of authority to the regulated firms and with the willingness to allow those firms to make more than normal profits in case of superior performance.

It is not just ideology and political climate or sector crises, due to inefficient regulation, that bring about sector reform. It is also technical, organizational and market developments that go along. The dynamics of technical change, convergence of markets and the explosion of demands occurring in the telecommunications sector overwhelm the traditional tools of regulation. They require, above all, competition that can only thrive

[1] See Averch and Johnson (1962) and Stigler and Friedland (1962).

under private ownership, the lack of legal entry barriers and extensive deregulation. Incentive regulation follows for whatever is left to regulate.

Before incentive regulation was developed, economists had derived and recommended optimal pricing formulas, starting with marginal cost pricing and leading to sophisticated Ramsey pricing formulas (that can be extended to nonlinear and peak-load pricing). The implicit assumption was that these formulas could be directly implemented, something that one of their discoverers, Marcel Boiteux, actually tried as an executive at Electricité de France. In contrast, incentive regulation acknowledges regulatory imperfections, moving from *optimal* regulation in the direction of *practical* regulation with desirable properties, something that was recognized early on by Baumol, who wanted to develop "reasonable rules" as a "plausible policy for an imperfect world" (Baumol, 1967).

The deregulation movement was both the result of adverse regulatory developments and the actual observation that markets could perform better (e.g., airline markets with competition in California and Texas). Similarly, incentive regulation was born out of the conviction that conventional regulation worked poorly and that better approaches, based on the profit motive could do better. As Crew and Kleindorfer (1996) rightfully observe, the introduction of incentive regulation in the U.K. was driven by technology and demand developments and "the iron will of Lady Thatcher." However, the specific solutions implemented were based on the economist Littlechild's (1983) suggestions that can be traced to Baumol (1967), Sudit (1979), Crew, Kleindorfer and Sudit (1979) and to Vogelsang and Finsinger (1979). It is not that practitioners could not have come up with these suggestions but that in this case they did not.

The Characteristics of Incentive Regulation

The modern incentive regulation literature differentiates between Bayesian and non-Bayesian regulatory mechanisms. The Bayesian mechanisms describe the regulator's lack of information by subjective probabilities that the regulator holds about parameters of the regulatory optimization problem. Based on the earlier principal-agent literature, Baron and Myerson (1982) developed the first Bayesian incentive regulation schemes in an adverse selection setting without cost information. Sappington (1983) added ex post observability of costs. Moral hazard was added by Laffont and Tirole (1986). The regulator is assumed to maximize his objective function under the constraints that regulated firms use their information advantage to maximize profits and that those firms are entitled to some minimum profit. Thus, the Bayesian mechanisms are derived as optimal, though in a restricted sense. Because of the subjective probability, regulatory mechanisms would depend on the person of the regulator and his or her state of information. Although the typical objective function postulated for the regulator gives less weight to profits than to consumer surplus and government revenues, a major result from the Bayesian literature is that, under the optimal mechanism, firms will make more than competitive profits. These profits are best raised in a fairly non-distortionary manner, for example, in the fixed fees of two-part tariffs.

Bayesian mechanisms have, through such insights, had a significant influence on regulatory economists and practitioners. They have, however, not been as much directly applied as the non-Bayesian mechanisms. It is thus not surprising that Sappington (2001), in his review of incentive regulation in telecommunications, hardly mentions the Bayesian approach that he so much helped to shape. The non-Bayesian mechanisms attempt to use only observable and verifiable (bookkeeping) data and to be independent of the particular regulator. Because those data cannot be foreseen, the mechanisms are not optimal.[2] What they strive for, is a stepwise improvement over the status quo, converging to an optimum over time. This process may, however, be frustrated by external changes to which the mechanisms can only react with a lag. The most popular incentive approach, price cap regulation, is a blend between Bayesian and non-Bayesian mechanisms. Its main Bayesian parameter is the X-factor, which has to be adjusted every few years. This adjustment clearly requires judgment about the potential of the firm to reduce its costs. Here is where the U.S. and U.K. approaches to price caps differ. The U.K. approach is more Bayesian in that it is trying to assess the future, taking market growth, new technical developments etc. into consideration, while the U.S. approach is more oriented toward past productivity growth rates.[3]

The incentive regulation approach led to several types of practical outcomes. The most important of these have been price caps, rate case moratoria, profit sharing, banded rate-of-return regulation, yardstick regulation, and menus. In end user regulation, price caps have become the most widespread. For example, in U.S. telecommunications the first state public utility commissions introduced price cap regulation in 1990, when already 26 states had some other form of incentive regulation.[4] However, by 2000, 40 states had price cap regulation and only 4 states had other forms of incentive regulation (Sappington, 2001). Similarly, the FCC moved from menu regulation of interstate access prices to pure price cap regulation. Most OECD countries and many other countries now use price cap regulation for their telecommunications sectors. In order to simplify the presentation I will only briefly characterize the most important types of incentive regulation and then concentrate on price caps.

Price caps are defined by an index of the regulated services that is adjusted, on an annual basis, by (1) an inflation factor that takes care of the economy-wide price level or of the level of input prices, (2) an X-factor that reflects efficiency improvements of the firm, and (3) a Y-factor that allows for pass-through of specific cost items outside the control of the firm. The index is further adjusted in regulatory proceedings over longer-term intervals of usually 4-5 years, taking into consideration the firm's performance under the price cap. This typically results in an adjustment of the X- and Y-factors. The use of a price cap index allows the firm some rebalancing of its price structure, where the

[2] An exception is the Loeb and Magat mechanism, which assumes that the regulator has full knowledge of the firm's demand, but is totally ignorant on costs. It is fully optimal if the regulator maximizes simply the sum of consumer surplus plus profits. See Loeb and Magat (1979).
[3] See Beesley and Littlechild (1989) and Crew and Kleindorfer (1996).
[4] Earlier, between 1980 and 1983, Michigan Bell was regulated by a form of price caps, but that was given up, when AT&T was divested.

incentives and opportunities depend on the index formula used. Empirically, the most relevant formulas have used chained Laspeyres indices or average revenue constraints.

Price caps, as we know them, were first suggested by Littlechild (1983) in a report for the British government on the regulation of British Telecom. They have, in my view, been so successful because they combine two characteristics that are essential for regulation today, (a) incentives for cost-reductions and (b) freedom and incentives for price rebalancing.

(a) The specific features of the first characteristic are that the cost-reducing incentives of price caps are fairly stable and viable. Both these features are important. In order to be high-powered, cost-reducing incentives have to hold over a long period of time.[5] In the long run, however, stability tends to interfere with viability, due to contingencies, such as business cycles, technical breakthroughs or structural changes in the economy. At some point in time, under stable prices, either the regulated firm goes bankrupt or it makes unconstrained monopoly profits, something that interferes with the regulatory goal of rent extraction and keeping customers happy. Price caps are viable, because they have built-in adjustments (in the RPI-X formula) that increase the regulatory commitment period and because they are subject to review every few years. The latter, of course, brings in ratcheting and has led some to equate them with rate-of-return regulation (Liston, 1993). Empirical work seems to contradict this assessment, though.[6]

(b) The second characteristic contributing to the success of price caps is the flexibility to change relative prices in the regulated basket of services, combined with a weighting scheme that promotes price rebalancing towards more efficient price structures. The use of baskets has provided the utility with the ability and incentive to rebalance their prices in the direction of allocatively superior prices and, possibly more important, has allowed regulated utilities to compete with new entrants.[7]

Rate moratoria can be viewed as the special case of price caps, where the Y-factor is zero and the X-factor equals the rate of inflation. Rate moratoria have strong incentive properties and are very popular with customers. However, they burden the utility with the inflation risk and can threaten its financial viability. Electric utilities and some telecommunications operators have felt this burden. As a result, some of these moratoria were followed by large adjustments that were viewed as rate shocks (Isaac, 1991). Today, rate moratoria have their place for specified basic services.

---

[5] The power, steepness or strength of cost-reducing incentives is usually measured by the percentage of cost reductions the regulated firm is allowed to keep.

[6] See Sappington (2001) and the literature cited there.

[7] See also Newbery (2000) for competitive properties of price caps. A very different view on the interaction between price caps and competition is taken by Lehman and Weisman (2000), who argue that regulators undermine the cost-reducing incentives of price caps by ex post setting more liberal terms of entry after first setting price caps. This could only be a one-time phenomenon. Its effect, if true, would be limited to the time between longer-term reviews.

Profit sharing or sliding scale regulation is probably the oldest regulatory incentive scheme, dating back to the 19th century in England.[8] It was also practiced for U.S. electric utilities in the first half of the 20th century. Profit sharing regulation lets the customers directly participate in excess profits or profit shortfalls earned by the utility. This can either take the form of ex post refunds or price reductions for future purchases. Profit sharing has strong fairness and quite good efficiency properties. It is, however, often administratively cumbersome, because profits have to be agreed upon every period. This becomes particularly difficult if only some of the firm's services are regulated.[9] Profit sharing was abolished in the U.S. in the 1950s, when affected electric utilities ran up losses. It then regained some popularity in U.S. telecommunications during the 1980s and 90s, but has been fading more recently.

Banded rate-of-return regulation lets the regulated utility keep its excess profits and suffer profit shortfalls within a pre-specified band. Only achieved rates of return below or above the band trigger rate cases to bring the utilities profits back inside the band. Banded rate-of-return regulation requires continuous monitoring of profits and is, thus, equally costly in administrative terms as profit sharing.

Yardstick regulation makes the prices the utility can charge dependent on the performance of other firms or on some efficient benchmark. Yardstick regulation is risky for a utility to the extent that its costs differ from the yardstick by virtue of such factors as geology, climate, population density, local wage rates, taxes, or the like. At the same time, yardsticks can provide strong incentives. They have been used successfully in cases, where cost data are not available, and in connection with other methods, such as for input price adjustments under price caps.[10] The main practical development, however, has been to base prices on efficient long-run costs, derived from micro-micro level engineering models of a utility. The idea for this can be traced to electricity regulation in Chile where, starting in the 1980s prices were based on long-run marginal costs of an efficient operator. The methods were refined in the U.S telecommunications industry, starting with cost models by Mitchell (1990) and leading to the so-called cost proxy models submitted to the FCC and other regulatory agencies in the U.S. These models try to measure the total long-run incremental cost of a service (TSLRIC, including fixed costs and a share of common costs with other services) or of a network element (TELRIC). The cost models used allow for an ultra-fine level of granularity and detail of telecommunications sector. Such detail is necessary in view of the sector's heterogeneity at company, country and service level. While these models try to include an enormous amount of local information, they will miss firm-specific peculiarities on input prices, demands etc. and therefore will not accurately measure the efficient costs of a specific telecommunications utility. They have, nevertheless gained widespread acceptance by telecommunications regulators around the world.[11]

---

[8] See Cave (2000) and the literature cited there.

[9] This was seemingly facilitated for FCC regulation of interstate access charges, because these are under cost separation from the LECs' other services. This, however, is only by accounting convention and does not measure true profits.

[10] For a broad set of examples, see Farrell and Mitchell (1998 and 1999).

[11] In the U.S., however, an appellate federal court has recently cast a shadow on their legality.

Menus or *options* allow the regulated utility a choice among different incentive regulation plans. This choice usually consists of combinations between price caps and profit sharing. The idea is to tailor the mechanism more closely to the specific circumstances of a utility without the regulator knowing beforehand, what the specific circumstances are. For example, the firm may have good or bad opportunities to cut costs. Because the choice made by the firm reveals its assessment, there is an ex post mechanism that the regulator (and usually the customers) prefer. In the same example, if the choice signals that the firm sees good prospects for cost reduction, a tighter mechanism would have been better. Menus therefore raise serious commitment problems. It is no surprise therefore that the FCC abandoned them after some time.

Price Caps and the Incentive for Cost Reduction

Price caps were invented in the early 1980s. They were first used by Michigan Bell during the 1980-83 period[12] and independently put forward in a report to the U.K. government by Littlechild (1983), leading to price cap regulation for British Telecom in 1984.

Price caps have been celebrated as a fairly straightforward way to provide incentives for cost reduction. This is an area, where traditional rate-of-return regulation and public enterprise pricing have been accused of failing. In the public-utility literature this failure is traced to the A-J effect, which showed theoretically, how rate-of-return regulation could lead to cost increases relative to an unconstrained profit-maximizing firm (Averch and Johnson, 1962). This discovery induced regulation economists to come up with suggestions to overcome such cost inefficiencies. It is thus here, where the greatest potential for improvement was seen. The potential for cost reduction through price caps stems from their property of simulating a long-term fixed-price contract. This was, in essence, Baumol's early contribution to the price-caps literature. He suggested institutionalizing a regulatory lag, during which prices would stay fixed (Baumol, 1967). It is, however, well known that fixed-price contracts only work for short periods. Then problems of breach, adjustment or renewal arise.[13] Baumol therefore picked three years as a reasonable, though arbitrary number. This suggestion is similar to a rate moratorium, with which it shares the lack of adjustment to changing circumstances. Adjustments to changing circumstances can help lengthen the regulatory lag but they have to be independent of the regulated firm's behavior.

The resulting inflation and productivity adjustments found in price caps go back to the discussion of fuel adjustment clauses (Baron and De Bondt, 1979) and to Sudit (1979) and Baumol (1982).[14] Sudit suggested price adjustments, based on the difference between a benchmark Divisia index of input price changes and benchmark changes in total factor productivities. He was very aware of the incentive properties as well as the adaptive

[12] For this see Brown, Einhorn and Vogelsang (1989) and the sources cited there.
[13] In the context of regulation, this problem has been analyzed by Goldberg (1976) and Williamson (1976).
[14] See also Crew, Kleindorfer and Sudit (1979).

properties of the pricing formula. In contrast to Sudit's input price adjustment, Baumol uses an adjustment for inflation. An inflation adjustment is definitely simpler and more understandable for the general population. At the same time, using an inflation adjustment more or less requires the X-factor to include both the productivity adjustment and an adjustment for the expected difference between the relevant input prices and inflation.[15] The failure to do so led to major disputes in the mid 1990s about the X-factor employed by the FCC for interstate access charges paid to local exchange companies.

Price caps seem to be more applicable, the smaller the Y-factor is relative to the X-factor. An extreme example have been the U.K. electricity supply companies in the early 1990s, for which the X-factor was small and the Y-factor large. In this case, the Y-factor covered more than 90% of the price.[16]

The beauty of price caps is that they finesse the fixed-price contract through the inflation adjustment *and* the X-factor (*and* Y-factor) *and* the pre-specified renewal date. This combination increases the commitment power but also decreases the cost-reducing incentives. As a result, the actual power of price-cap incentives may not exceed that of traditional regulation by a large amount. This then becomes an empirical issue.

Price Cap and the Incentive for Allocative Efficiency

The question of allocatively inefficient pricing under rate-of-return regulation can also be connected to the Averch-Johnson literature, which showed that rate-of-return regulation could lead to price distortions, such as the suppression of allocatively efficient peak-load pricing. It was further influenced by the literature on sustainability and cross subsidization (in the aftermath of Faulhaber, 1975) that showed the lack of sustainability of (cross-)subsidized price structures. The similarity of profit-maximizing and welfare-maximizing price structures led to the proposal by Vogelsang and Finsinger (1979) to let utilities choose their own price structures, as long as they maintain a certain price level for the basket of their services. Allocative efficiency is much less tangible than productive efficiency. It may also conflict with fairness or distributional concern. As a result, the allocative efficiency properties of price caps were initially viewed by many as less important than the productive efficiency properties. At the same time, issues of cross subsidization have traditionally been as important for regulators as average prices charged by regulated firms and as productive efficiency.

The standard non-incentivized regulatory price structure is known as fully distributed cost pricing. It is related to rate-of-return regulation, because it also tries to make prices fully dependent on costs and on traditional methods of accounting. As has been shown by several authors, fully distributed cost pricing is compatible with any arbitrary price structure (Braeutigam, 1980, Baumol, Koehn and Willig, 1987). It can therefore be used by regulators and pressure groups to generate desired cross subsidization. Fully-

[15] These relationships are very clearly worked out by Bernstein and Sappington (1999).
[16] See Armstrong, Cowan and Vickers (1994).

distributed cost pricing, as applied by regulators, is usually not compatible with efficient pricing and with effective competition. In order to have these two desirable properties, price structures require some flexibility plus constraints or incentives to use this flexibility in the right way. This can be done with several methods. First, the flexibility can be constrained by upper and lower bounds, such as (benchmarked) stand-alone costs and incremental costs or simple percentage deviations from current prices. The former method is applied by Oftel to BT's access charge price caps, while the latter method is a standard feature of FCC price caps. Second, optional prices can be allowed in addition to regulated prices. Third, the regulated utility's services can be placed in baskets within which prices can be rebalanced subject to a constraint on their level. This level can be described by a price index (usually Laspeyres) or by average revenue. In line with our emphasis on price caps, we will consider average revenue constraints and different weights for price cap indices. Flexibility of price structures then depends on the type of index and the scope of baskets.

Littlechild (1983) based his price caps on a chained Laspeyres price index. Such price caps have the desirable theoretical property that, under stationarity assumptions, prices converge to Ramsey prices.[17] This result, however, fails if demands and costs change too quickly (Neu, 1993, Fraser, 1995). The improvement over time may therefore turn out to be absent or insufficient. Idealized price-cap weights would lead to Ramsey prices right away and therefore not be subject to lags or strategic behavior. Such weights equal correctly forecasted optimal quantities (Laffont and Tirole, 1996). Since such forecasts require the same (forbidding) amount of information as the derivation of Ramsey prices, the question arises if better approximations exist than the quantities of the last period. Several combinations of current and last period's quantities have been suggested for this purpose, including a simple average between a Laspeyres and Paasche index (Vogelsang, 1988) and the Fisher ideal price index (Diewert and Fox, 2000). Such averages could considerably speed up convergence and thereby immediately improve the regulatory outcome. They could, however, be subject to strategic manipulation (Vogelsang, 1988) and can only be calculated ex post (so that ex post price adjustments are necessary).[18]

Price caps for some regulated sectors in the U.K. contain average revenue constraints rather than price indices. Such constraints, which divide total revenues by total quantities, allow for more flexibility and appear to be simpler than price indices. They suffer, however, from at least four shortcomings. First, they can be subject to manipulation (Sappington and Sibley, 1992). Second, they tend to lead to bad welfare outcomes (Bradley and Price, 1991; Armstong, Cowan and Vickers, 1994). Third, they may require

[17] See Bradley and Price (1988) and Vogelsang (1989). The result is based on Vogelsang and Finsinger (1979). The basket approach also has the advantage that it lets the firm do rebalancing that regulators and politicians want but would like to avoid to be blamed for. However, regulated firms also like to avoid such blame. For example, Deutsche Telekom keeps its monthly fees for telephone service below loop costs, although it is allowed to raise them.

[18] Paasche indices have very undesirable properties and, to my knowledge, are not directly applied anywhere.

ex post adjustments. Fourth, they require homogeneous commodities. However, with all these shortcomings, they are very simple and flexible.[19]

Like the productive efficiency properties, the allocative efficiency properties of price caps, while potentially impressive, are compromised by practicality. Idealized weights, as suggested by Laffont and Tirole (1993), are not implementable in practice. Most practical Laspeyres weights, however, only converge to efficient prices under unrealistic conditions. Again, the actual efficiency properties have to be established in empirical work.

Price Caps and Competition

From the beginning, one of the main goals pursued with the introduction of price caps has been to accommodate competition. It was the presence of competition that had made rate-of-return regulation obsolete and triggered the FCC and the British government to look for alternative ways of utility regulation. Nevertheless, Littlechild had put major emphasis on price caps as a tool to reduce monopoly power. He did not seem to anticipate that possibly the main property of price caps as a successful regulatory mechanism would be their strong survival properties in a competitive environment. Littlechild had wanted to introduce price caps for monopoly services only. That's why he termed them a "local tariff reduction scheme." However, he already saw part of their competitive compatibility in that they would allow a set of competitive outputs of a regulated firm to be separated from the monopolistic outputs. Because price caps do not require measuring the overall profitability of the firm, they can be restricted to subset of the firm's outputs.

In order to assess the competitive role of price caps, we first look at the problem of horizontal market power, cast as the natural monopoly issue. In the next section, we then turn to vertical issues that have, over the last 20 years, gained center stage.

Natural monopoly has traditionally been the main justification for public utility regulation.[20] Twenty years ago, natural monopoly was assumed to be present in most such industries. In telecommunications, natural monopoly was doubted only for long-distance markets but clearly assumed to hold for local telephony. The questions hotly discussed then were if natural monopoly could be either unsustainable or contestable. Today, these questions have faded. It has become unclear to what extent natural monopoly prevails at all. Unsustainability is only perceived as a problem in connection with cross subsidies imposed by regulators, and contestability is only viewed as having a chance if entrants are helped by access and interconnection regulation. The power of potential competition has been replaced by the power of actual competition. The main question, when moving from a regulated or state-owned monopoly to competition, is if the former monopolist has sufficient advantages over entrants to be able to stifle

[19] For an attempt to correct some of the shortcomings of average revenue constraints see Vogelsang (forthcoming).
[20] In telecommunications and electricity transmission, externalities also play a major role.

competition. The answer depends largely on the presence and significance of sunk costs (and on network interconnection, discussed in the next section). In telecommunications, most analysts concentrate on the local loop as the basis for both natural monopoly and sunk costs. The local loop has been remarkably stable and long-lived. It is also the most costly item in the network. However, views differ on the amount of technical and market changes that would reduce the incidence of sunk costs.[21]

With such differences in the assessment of the market conditions, what consequences should one draw? One consequence is certainly that more empirical research is needed. Again, consider the case of telecommunications. Here the econometric estimates of cost relationships show that the evidence for natural monopoly in the local network is weak, while, at the same time, proxy cost models come to the conclusion that local network costs are a sometimes steeply declining function of network density. Thus, there appears to be a discrepancy between the two approaches at least for networks of low to intermediate density.[22] The question is, which models are correct? This would be of prime importance for the policy consequences. If the econometric models were correct, there would be a strong case for pursuing infrastructure competition throughout the network. If the engineering models were correct, infrastructure competition in local networks would make sense only in very dense networks, where economies of density are exhausted. In other areas, competition would be restricted to some form of resale (of services or of network elements). There is, however, a different policy consequence that might be drawn at the current stage of our knowledge. It is that we have to find a policy approach that makes us less dependent on the research outcome about the presence of natural monopoly, economies of scope and scale and sunk costs. This policy approach starts with the observation that, conceptually, there could exist two important ranges of natural monopoly.

In the first range, the natural monopoly property is *weak*. In this range, economies of scale and scope are almost exhausted and sunk costs tend to be small. If we have this situation, which is usually characterized by dense networks with many users, competition is likely to be beneficial, because it leads to pressure on costs, prices and innovation. Competition is also likely to occur to the extent that demand is moderately to strongly inelastic. Thus, with free entry we can expect duplicate network investments, associated with some cost inefficiency and excess capacity but possibly lower prices than under regulated monopoly.

In the second range, the natural monopoly property is *strong*. In this range, there are clear economies of scale (and scope) and sunk costs tend to be large. Examples include telephone access networks in rural areas and electricity transmission networks in general. In such situations, competition would most likely be wasteful and not self-sustaining. However, if we allow it and do not subsidize competition it is unlikely to come.

---

[21] See, for example the different views of Spulber (2001) and Woroch (2001).

[22] See Fuss and Waverman (2001) on the econometric estimates and their problems and Wimmer and Rosston (2000) and Sharkey (2001) on engineering cost models.

We can now do an assessment of the type I and type II errors arising from different entry policies (initially assuming no regulation). First, consider allowing entry, looking at three possible situations but not knowing which one prevails. The first is no natural monopoly at all. Here, entry is clearly called for. So, the error from denying entry is likely to be large (although, there is a slight probability that entry may be excessive and, at the same time, the loss from this greater than the loss from unnatural monopoly). Next, consider a weak natural monopoly. If we allow it, entry may or may not occur here. If it does not occur no harm is done. If it does occur, the potential increase in resource costs needs to be balanced against the allocative and cost-reducing benefits from competition. Thus, if we allow entry in a weak natural monopoly there is some likelihood of a net gain and some likelihood of a small or moderate net loss. The expected value of this is likely to be non-negative. Last, consider a strong natural monopoly. If we allow entry here, it is quite unlikely to occur, because strong natural monopoly is usually associated with entry barriers such as large sunk costs. If it occurs, however, there is likely to be a major loss of resources, due to duplication. Considering the same three cases for a policy of blocked entry, we see that there is likely to be a major loss for the case of no natural monopoly and a major (though fairly unlikely) gain in the case of a strong natural monopoly. The conclusion of all this is that, absent regulation, entry barriers should be abolished, unless there is clearly a strong natural monopoly.

Now, add to this price-cap regulation that keeps the incumbent utility's prices at fairly efficient levels with a fairly efficient structure. My conjecture is that such pricing would eliminate any dangers from opening the market because the prices, in case of a strong natural monopoly, would already be so low that an entrant could not break even. Under price cap regulation with some form of commitment, entrants would expect that prices could not be raised, due to entry. As a result, with strong natural monopoly, entrants could not hope for the incumbent to increase prices to recover the cost increase from entry (this would take the form of thinning out the subscriber base of a network). In contrast, under rate-of-return regulation, prices could be increased, provided demands are sufficiently inelastic.

Thus, independent of the potential presence of natural monopoly, there is little harm in allowing competition in public utility industries in general, and the remaining harm might be eliminated with price caps. This consequence is, of course, well known at this time and practiced in wide parts of the world. There is, however, another less-appreciated consequence. It is that our yardstick for measuring the competitiveness of telecommunications markets should be the amount of actual competition that occurs if markets are fully opened. This, in a way, replaces the role of natural monopoly analysis. Rather than asking the question, whether a market is a natural monopoly, we ask, what amount of competition occurs. If there is no competition (in spite of open markets), we may suspect the strong version of natural monopoly to hold. Whether or not that is the case, such a market requires, in our view, further regulation. If competition occurs and spreads out quickly, we may suspect the absence of natural monopoly or the presence of only weak natural monopoly. In this case, deregulation is warranted. If some competition occurs but the incumbent former monopolist remains dominant, we have the intermediate case, where relaxed regulation may be in order, but not yet full deregulation. Competition

cum regulation thus becomes the discovery process for the existence and strength of natural monopoly and for guidance on the policy of deregulation. In order to fulfill this function, regulation has to be made competitively neutral. This requires three tools:

- The abolishment of legal entry barriers,
- A functioning incentive scheme for end-user regulation,
- A competitively neutral scheme of access regulation.

Since the last of these requirements concerns vertical relationships, we postpone its discussion to the next section. Here, we concentrate on the second, assuming that the first is already accomplished. We use price caps as the example.

To simplify the analysis, we consider only two dimensions, the *tightness* of the price cap, as expressed by the X-factor, and the *flexibility* of the price structure, as expressed by the size of the basket. Price caps are compatible with competition because competition requires flexibility and because it constrains rents and therefore makes profit control less important. However, competition and price-cap regulation are potentially in conflict, when the flexibility is used to chasten competitors or when price caps are so tight that competitors are kept out.

In bringing out the properties of tightness, we assume a single-product firm. On average, the tighter the cap, the lower will be the utility's prices and the less entry is attracted. Thus, we can expect a negative correlation between tightness of the cap and the amount of competition (Abel, 1999).

Can a cap be too tight, in the sense that it prevents entry that would have been desirable? My conjecture is that it cannot, as long as the regulated utility remains viable. In this case, lack of entry would signal the presence of at least weak natural monopoly. Couldn't there be other entry barriers that, along with the low price, would prevent entry in unnatural monopoly so that higher prices and entry would be preferred? If this were the case, entry could not lead to lower prices and thus would not be beneficial. Also, entry would not likely occur at slightly higher prices, because entry barriers are defined by the ability of the incumbent to keep prices above costs without incurring entry.

Now, consider price caps that are too loose, in the sense that excessive entry occurs. This could happen under all three natural market structures discussed above. It is least likely under strong natural monopoly because there prices would have to be indeed high to attract entry. It could, however, occur under weak natural monopoly and under natural oligopoly. In both of these, we would, in the worst case, likely get the same result as without regulation. As shown in the literature, excessive entry can have some efficiency costs. This literature is, to my knowledge, largely restricted to the weak natural monopoly case (assuming some fixed costs and constant variable costs).[23] Here, the potential efficiency losses are largest. They could, however, also turn up in natural oligopoly (under U-shaped cost curves). In both cases, tighter price caps could cure the excess entry

[23] For a lucid exposition, see Armstrong, Cowan and Vickers (1994).

problem.[24] Thus, relative to the right amount of competition the potential error from too loose price caps appears greater than the potential error from too tight price caps. Obviously, however, from the point of view of the viability of the regulated utility and partially from the point of view of cost-reducing (and investment) incentives, the assessment is reversed. The tightness of actual price caps therefore has to be a compromise.

Now, consider the effect of flexibility of the price structure of a multi-product utility. To simplify, assume the firm operates in two markets, one with strong natural monopoly property and the other a natural oligopoly. No flexibility in the price structure here means that each has its own price cap. Thus, the analysis of the last paragraph largely carries over (except that there could be economies of scope and/or cross-elasticities of demand). The results could change, however, if one assumes flexibility, by having a single price-cap basket for both markets. The utility could then reduce the price in the potentially competitive market, thereby keeping competitors out, and simultaneously increase the price in the strong natural monopoly market (without attracting competitors there). This will be profitable if the loss in the potentially competitive market is smaller than the gain in the strong natural monopoly market. Abstracting from the effects of weights of the price-cap index, the profitability of such a move likely depends on the tightness of the price-cap constraint and on the prior price structure. To take one extreme, if the constraint is not binding there is no gain from increasing the price in the strong natural monopoly market that could be used to deter entry in the potentially competitive market. Thus, entry would occur in the natural oligopoly market.

What happens if the price-cap constraint is binding? To simplify, assume Cournot competition (or differentiated Bertrand competition, so that the regulated firm's profit function is differentiable) and start at the unconstrained Cournot oligopoly outcome in market 2. That means, the regulated utility's profit function is flat at this point. In contrast, the price in market 1 is constrained by regulation. Hence, the firm's profit function is increasing in this price. Thus, by the envelope theorem, the firm can increase its profits by raising the price in market 1 and lowering the price in market 2. Thus, a small reduction in price in the natural oligopoly market would lead to a second-order loss, while the resulting increase in the price of the strong natural monopoly market would lead to a first-order gain.[25]

As another way to see this, consider a Laspeyres price cap index for the two markets. It is easy to see that the desired price changes of the regulated utility depend on its current markups and the price elasticities. In fact, profitable price changes are described by a move in the direction of the Ramsey pricing rule, with the qualification that the relevant elasticities are firm specific. Under independent demands, this means that the firm will increase the price in the strong natural monopoly market and decrease the price in the

[24] Natural oligopoly with flat-bottomed average cost curves would also suffer no efficiency loss from tight price caps.

[25] The profit function may not be differentiable with respect to a price change in the potentially competitive market. This could occur if a price increase triggers discrete entry. Such discontinuity would likely work in the same direction as the argument above.

potentially competitive market if markup times elasticity in the former is smaller than markup times elasticity in the latter. In the absence of changes in demands, in technologies and in input prices the sequence of prices will converge to those Ramsey prices.[26] The question is what their welfare properties are relative to conventional Ramsey prices. The candidate welfare optimum to compare them to is known as the *autarkic* Ramsey optimum (Baumol, Panzar and Willig, 1982, p. 344), which is defined by all firms in the markets independently striving to maximize welfare, subject to breaking even. Such an autarkic Ramsey optimum would be the best to hope for under price caps with competition.[27] Usually, the outcome will be worse, because (a) the regulated utility will make a positive profit and (b) the other firms do not maximize welfare. In any case, the above argument on first-order gain and second-order loss is restricted to prices that were not Ramsey optimal in this sense. While not Ramsey optimal, the resulting equilibrium prices will improve welfare over unconstrained monopoly prices and welfare will increase in the tightness of the constraint, provided the regulated firm's viability is assured. We would also not expect that the incumbent utility would charge a price below marginal cost in the potentially competitive market.

Without looking at dynamic incentives for strategic anticompetitive behavior, we have established the conjecture that tightness and flexibility of price caps provide (myopic) incentives to improve upon the unregulated outcome in the potentially competitive market. Unless it can be shown that the myopic incentives are undermined by contrary strategic incentives, this conjecture would make a case against restrictions on rebalancing, such as the bands imposed upon the local exchange companies (LECs) by the FCC. Such bands could still be justified as a tool for improving fairness, because the economic cost of rebalancing is felt mostly in the strong natural monopoly market. Assuming that the advent of competition makes the firm-specific demands of the competitive goods more elastic, such rebalancing would also happen under deregulation of the potentially competitive markets, but less than under price caps. Even under price caps the price in the potentially competitive market will remain above marginal costs.[28] So, no inefficiency arises here. In fact, efficiency is likely improved.

The proviso that there is no strategic behavior is of course an important one. What importance could strategic behavior gain in the discussed regulatory setting? Since lowering the price in the competitive market can partially be compensated by a price rise in the captive market, *predatory pricing* could be a possibility. This would, however, only be likely, if the incumbent could successfully increase profits after successfully eliminating the rivals. Since the average price in both markets would remain unaffected by exit of the rival, the profit increase would have to come from a change in price structure. The exit of rivals could have two effects here. First, it could reduce the incumbent's price elasticity in the potentially competitive market and, second, the

[26] See also the discussion of global price caps below in the section on vertical relationships.
[27] In welfare terms, the autarkic Ramsey optimum is dominated by the viable firm Ramsey optimum, (which allows for coordination among firms) which in turn is dominated by the viable industry Ramsey optimum (which allows for side-payments among firms, in order to meet a common budget constraint).
[28] Unless predation or other anticompetitive behavior occurs and unless complementarity between market demands of the outputs is very strong.

incumbent's demand function would shift outward. As a result, price and profit in this market would increase, only partially compensated by a price and profit decrease in the captive market. Predatory pricing is thus a possibility, provided the entrants face reentry barriers.[29] The second possibility for strategic behavior would be a credible threat that the incumbent would increase output in the potentially competitive market if entrants increase their output. Since this is milder than full predation, it is quite credible. As a result, there could be a collusive equilibrium, where the entrants hold back output and the incumbent therefore enjoys a larger output. Strategic behavior could be influenced by additional regulatory factors, the most important being the potential lack of regulatory commitment. For example, regulators cannot easily commit to deregulate (or not to deregulate). If the regulated utility wants deregulation and if deregulation depends on the demonstration of a large market share of rivals, then the utility will keep its prices in the potentially competitive market high and, at the same time, please the regulator by keeping the price in the strong natural monopoly market low. In contrast, if the utility wants to perpetuate regulation, it will keep rivals out by lowering prices in the potentially competitive market. Strategic cross subsidization, while not directly profitable, can help maintain the regulated monopoly, because it acts as an entry deterrent in the market with flat costs without necessarily attracting cream skimming in the strong natural monopoly market. Overall, strategic anticompetitive pricing appears to be possible under flexible price structure, and should therefore be guarded against.[30] This is a potential weakness of relying on general rules, because such guarding may require cost measurement.

Regulation of Vertical Relationships

Competition substitutes for regulation and improves on it by putting pressure on regulated firms to reduce their costs and to align their prices with costs. In that sense, competition facilitates regulation. At the same time, the relationship between the regulated utility and its competitors now becomes part of the regulator's concern. Thus, while potentially improving on the outcome, competition makes regulation harder to do.

For the last few years, the decline of *legal monopolies* has been a worldwide phenomenon. Competition (often accompanied by privatization) has been introduced in traditional monopoly sectors, such as telecommunications and electricity supply. There has been a similar decline in the reported incidence of *natural monopoly* in these sectors. Only niches of natural and, in some cases, legal monopolies remain. Examples could include local telephone companies, gas distribution companies, electricity distribution and transmission companies. In spite of this decline, why might it nevertheless be worthwhile to cast a new look at the regulation of these monopolies in light of limits to optimal regulation? All the above industries have potentially competitive parts that are vertically related to monopolies. In the energy sector, electricity transmission and distribution have strong monopoly attributes, while generation and marketing are

[29] To the extent that entry barriers take the form of sunk costs, rivals are less likely to exit, because their assets remain.

[30] Under rate-of-return regulation, anti-competitive behavior can be the result of cost shifting and A-J effects.

potentially competitive. Natural gas and oil pipelines and gas distribution companies are potential monopolies, while gas and oil production could be competitive.

Monopoly regulation today almost inevitably has to be seen in the context of related competitive industries. Furthermore, the monopolists often are vertically integrated so that they either hold a monopoly over all stages or compete with firms that use the monopoly outputs as their inputs. Thus, two important questions for public utility regulation today are, 'What should be the vertical structure of the industry?' and, given the vertical structure, 'How should the outputs of the dominant utility be priced?' The answers to the two questions are interrelated, because the answer to the first depends on the imperfections in pricing. One reason for a vertical breakup of the dominant utility can be that vertical integration interferes with pricing of competitors' access to its essential facilities. Since we are here interested in the pricing aspects only, we separately consider pricing under a vertically integrated and under a vertically separated scenario. The main differences between vertical integration and vertical separation in this context are

(a) that the vertically integrated utility competes with the buyers of its intermediate inputs, while the separated utility only sells at one level. As a result, the integrated utility can have incentives to disadvantage those buyers of intermediate inputs.
(b) that the vertically integrated utility may enjoy economies of scope that the vertically separated utility may not. As a result, vertical integration may reduce total costs of the industry.

Consider first the case of *vertical separation* into monopolistic and competitive stages. Here it is clear that the monopolistic stages would be regulated. This is essentially the horizontal case analyzed in the previous sections on the monopoly case. The competitive stages are less clear-cut. Their regulatory treatment depends on the extent to which the former integrated monopoly was split up horizontally at a competitive supply stage. If no horizontal divestiture occurred, the successor company at this stage may start out as a monopoly and may therefore require regulation, presumably in the form of price caps. Because of the purchase of essential inputs from a regulated monopoly stage, those input prices would be combined in the Y-factor of the price-cap index. Otherwise, price caps would work as elaborated in the previous section. In particular, this regulation could help discover the competitive properties of this stage. The vertically separated case is thus, for the most part, a set of horizontal cases.

The case of *vertical integration* is much more complex. For simplicity, we assume two stages, where the first stage provides an essential input, available only from the vertically integrated regulated utility. The utility and potential rivals compete in the second stage for end users.[31] In terms of price caps, we can have, (1) the essential input regulated by TELRIC/TSLRIC, combined with (a) price caps for end users or (b) deregulation of end-

[31] We are not considered two-way connection issues, which play a major and increasing role in telecommunications and the Internet. Here, collusive issues arise because, under certain circumstances, the reciprocal access charges can function as a perfect (and almost costless) instrument of coordination.

user services, (2) price caps for the essential input and for end users, which could be (a) separate or (b) global price caps for both stages.

Case 1: Input regulation via TELRIC/TSLRIC

We have discussed the benchmarking properties of input regulation via TELRIC/TSLRIC above. In contrast, the main concern of this section is with the competitive properties of keeping the prices for first-stage essential inputs at TELRIC/TSLRIC levels. In order to abstract from measurement problems, we first assume that the prices for essential inputs are held exactly at the level of average incremental costs (plus appropriate share of common costs). Furthermore, end-user prices are regulated in such a way that the regulated utility would make a competitive return as a monopolist.[32] The essential inputs are characterized by strong natural monopoly, and there are economies of scope for the utility between the essential inputs and the end-user service. The end-user service can be produced by potential entrants in a weak natural monopoly or natural oligopoly setting, without economies of scope.

In the assumed vertical context the concepts of incremental and stand-alone costs and therefore the concept of economies of scope are not easily defined. The problem is that the essential input is always part of the final output. Thus, there is no incremental cost of the essential input, given the production of the final output. Perhaps this is the reason why in practice the measurement of TELRIC/TSLRIC (in the bottom-up approach) usually proceeds as if the network that is considered the essential input is treated as a stand-alone network. In our analysis, we assume that a non-integrated rival has costs that are additive in the essential input stage and the final output stage. In contrast, the incumbent is assumed to use a common input to produce the essential input and the final output. Economies of scope here mean that a vertically integrated efficient incumbent can produce the final output at lower costs than the sum of a non-integrated producer of the essential input and a producer of the final output. Economies of scope then can either mean that the incumbent has lower variable costs in the second stage than the rivals or that the sum of fixed costs for two vertically separated firms for the two stages is smaller than the common fixed cost for an integrated firm. Assuming fixed proportions, a vertically integrated firm of the latter type would have costs $C = K + c_1Q + c_2Q$, while a stage 1 firm would have costs $C_1 = K_1 + c_1Q$ and a stage 2 firm $C_2 = K_2 + aQ + c_2Q$, where a is the access charge for a stage 2 firm. Economies of scope are captured by $K < K_1 + K_2$. The "incremental" cost of stage 1 would then be defined as $IC_1 = K – K_2 + c_1Q$, which has similarities with the efficient component pricing rule (ECPR).[33]

We want to answer three questions. First, can the incumbent survive? Second, to what extent is TELRIC/TSLRIC an entry help? Third, what are the overall efficiency properties?

---

[32] We are not here looking at the potential stranding of the incumbent's assets, due to lack of past depreciation or due to low estimates of TELRIC/TSLRIC or lack of regulatory commitment. We also neglect the effects of short-run fluctuations in costs and demands.

[33] The extension to multiple outputs and multiple essential inputs is conceptually straightforward but would be difficult to implement.

1. Consider the first question. Since the incumbent breaks even as a monopolist and the price caps are binding, entry can only occur if entrants have lower costs than the incumbent. To the extent that the incumbent sells the essential input(s) at incremental costs, he passes on all the advantages from economies of scope to the entrants. The share in common costs, however, reduces this pass-on. Thus, lacking the incumbent's economies of scope, an entrant could only have lower costs than the incumbent if he was more efficient. In fact (assuming the absence of *dis*economies of scope), the entrant must have a cost function that is additive in the essential input and the second production stage. To the extent that entrants and the incumbent experience economies of scale from the production stage for the final good, the price caps would keep entrants out that are only mildly more efficient.[34] Thus, the combination of price cap and TELRIC/TSLRIC keeps an efficient incumbent viable and does not put the incumbent in a competitive disadvantage. However, entry does pose a problem for a slightly inefficient incumbent, because the replacement of end-user sales of the incumbent associated with entry leads to an endogenous increase in costs per unit of final output sold by the incumbent. Assume the extreme case, where the incumbent's end-user sales drop to zero and are totally replaced by those of entrants, who buy the essential input(s) at incremental costs plus a share in common costs. The survival of the incumbent, in this case, hinges on the common cost allocation rule. If common costs are allocated by the relative quantities sold the incumbent will always stay viable because, as he loses end-user market share, the essential input prices would move from incremental costs to stand-alone costs.[35] This move is also in line with the notion that (for a given cost function) the optimal access price is an increasing function of the intensity of downstream competition. That is, if downstream competition is weak, access charges should be low. This has the effect of compensating for downstream allocative inefficiency and of increasing downstream competition, while contributions to common costs are collected downstream. If downstream competition is fierce no markups are collected there and therefore have to be collected in the markets for essential inputs. Furthermore, if upstream bypass is possible, access charges at cost improve overall cost efficiency.

2. Now, turn to the second question. TELRIC/TSLRIC is an entry help to the extent that entrants benefit from the incumbent's economies of scope. However, if price caps for end user services are stringent, entrants only benefit if they are more efficient than the incumbent.

3. Now turn to the third question. In what sense is overall efficiency increased? If the entrants can provide stage 2 of the end-user service at lower costs than the incumbent, total costs are reduced. The overall result is thus likely to be productively efficient. Bypass at stage 1 occurs if an entrant's costs of bypass are less than or equal the benchmark incremental cost of access. Entrants compete downstream definitely if they

[34] To see this, just assume that an entrant captures 50% of the market and that, in order to do that, price decreases by some small amount. Then, each firm's output will be less than the monopolist's output was before and, because of economies of scope, average costs will be higher. Thus, entry is unprofitable.

[35] However, Armstrong (2001) shows that the ECPR is optimal in this context if the downstream price caps use idealized weights that are proportional to the (autarkic Ramsey) optimal outputs.

are productively more efficient downstream. A less efficient entrant can compete only if the incumbent has a markup downstream that exceeds the cost difference.

The allocative efficiency of the outcome will depend on the way price competition and end-user regulation work. As indicated above, the incumbent utility may keep end-user prices strategically low in markets with potentially much competition and high in markets with little competition. The common cost allocation weakens the productive efficiency results and can interfere with allocative efficiency, since a proportional markup conflicts with the Ramsey rule.[36] It may be below or above the corresponding Ramsey price and thus may distort allocation and the incentives of the incumbent. For example, if the access price is below the corresponding Ramsey price the incumbent may have incentives to deteriorate the quality of access. This could trigger specific quality regulation, something that has increased in importance under incentive regulation with binding constraints. Parity rules, which require an incumbent to provide essential facilities to its rivals at the same quality as to itself, are a potentially good benchmark. However, definition and implementation of parity rules are difficult.[37]

Deregulation of end-user prices is the natural consequence if downstream competition has sufficiently developed and if price caps are no longer binding. In fact, if price caps are binding one would underestimate the amount of competition that would result under deregulation of end-user prices. As long as TELRIC/TSLRIC regulation of essential facilities persist, competition should then remain viable.[38]

All this requires a major caveat. We have assumed that cost measurement was exact. As explained in our initial remarks about yardstick regulation, one of the problems of yardsticks (such as TELRIC/TSLRIC) is that they can deviate substantially from the efficient costs of an operator. Omitted variables, wrong measurements, geographic averaging, time fluctuations or other reasons can cause TELRIC/TSLRIC to deviate from the efficient costs of an incumbent. Since these errors affect prices for essential inputs, they can result in serious cream-skimming problems. Whenever the relevant efficient costs of an incumbent are lower than TELRIC/TSLRIC, the incumbent gains a competitive advantage. Conversely, the incumbent is disadvantaged, whenever the efficient costs are higher than TELRIC/TSLRIC. Because of these errors it is important that cost models are developed and implemented in an open process that avoids biases.

Why has the TELRIC/TSLRIC methodology become so widespread in markets for essential telecommunications facilities? There are several reasons that together provide a fairly strong explanation. First, for many of these facilities no markets and therefore no prices existed prior to the opening of the end-user markets. Thus, regulated prices had to be created. Creating price caps using zero starting quantities would, for example, leave

[36] As mentioned above, the way TELRIC/TSLRIC are sometimes calculated, they are closer to stand-alone costs than to incremental costs.
[37] See Farrell and Mitchell (1999).
[38] We do not treat deregulating the prices of essential inputs, while end-user prices are regulated. That seems to have it the wrong way, because the monopoly is deregulated, while the competitive part remains under regulation.

too much discretion to the regulated incumbent. Second, the essential facilities have fairly well-defined engineering properties that lend themselves to quantitative modeling. Third, the buyers are experts that can assess the validity of the models and can monitor quality of service. A fourth reason is that such prices look advantageous to entrants and end-users, both of which have influence on regulators. The fifth main reason is their ability to provide rather tailor-made prices for specific items at a specific location. The sixth is that they provide the right signals to rivals for bypassing the essential facility. The seventh is their strong incentive property, based on the absence of a link between the potentially inefficient cost of a regulated incumbent and the prices under the TELRIC/TSLRIC methodology. Some economists argue that the efficient firm standard can subvert the incentives to innovate, because that then becomes the new efficient standard. While it is true that new technologies and new practice influence TELRIC/TSLRIC, these efficient costs are mostly influenced by industry-wide paradigm changes that are rarely influenced by individual firms. Also, the lags involved are quite long. What is more of a concern is that the models are complex and not easily updated so that TELRIC/TSLRIC measurements can act as input rate moratoria.

Case 2: Price Caps for Access and End Users

The disadvantages of TELRIC/TSLRIC regulation probably gain out over their advantages, as time and competition progress. It is therefore natural to look for other methods for input price regulation, price caps in particular. Price caps could maintain cost-reducing incentives and improve on the allocative efficiency. Under TELRIC/TSLRIC pricing of essential inputs a price squeeze by the incumbent could not emerge. In contrast, under price caps a price squeeze cannot be ruled out. Having separate price caps or separate baskets for end-user services and for essential inputs largely eliminates the danger of a price squeeze or other strategic behavior. At the same time, separate price caps impose rigidity regarding rebalancing. This rigidity could be avoided by global price caps.

Global price caps were suggested by Laffont and Tirole (1996, in the following: L-T). They apply a single price-cap index to both, end-user services and to the essential inputs provided to competitors. L-T choose idealized weights based on Ramsey-optimal quantities. Interpreting input demands as derived demands, they conjecture that global price caps lead to Ramsey prices (in the sense of the autarkic Ramsey optimum discussed above). They are confident that anticompetitive incentives under global price caps are minimal but nevertheless suggest an imputation rule as a safeguard. The imputation rule says that the access charges have to be weakly below the (ex post) efficient component price. The use of idealized weights makes this constraint innocuous, unless the Ramsey prices would violate the efficient component-pricing rule (ECPR).

As stated above, however, idealized weights, in practice, have to be replaced by realistic weights, such as chained Laspyres weights. The question is if we can still expect the global price caps rule to have the desired properties. First, do we need an imputation rule? We conjectured above that such strategic behavior could emerge under horizontal competition. What is different now is that the captive market is for the essential input.

From the incumbent's point of view, selling to entrants and selling to final users are substitutes. With a binding price cap, an increase in the access price automatically leads to a decrease in the price for end users. Assume that we start at the efficient component price, with the incumbents and rivals breaking even and both having the same per unit costs in the downstream stage. Would the incumbent want to decrease the end user price? That would trigger an increase in the access price, squeezing out the rivals. The incumbent's total margin per unit would thus fall, although its total sales would increase. Profits could fall, rise or stay the same. Squeezing is thus a possibility. What about predation? Assume that squeezing is not directly profitable. It can nevertheless induce rivals to exit at low costs to the incumbent. Provided, the rivals do not reenter, the incumbent could then reduce the access price and increase the price for end users and make large profits there. However, this possibility strongly depends on the entry barriers for rivals. Since the essential input is now provided at a low price, any other entry barriers could be overcome more easily. Thus, the price caps have a built-in insurance against predation.

Conclusions

Over the last 20 years, public utility regulation has found new tools in the form of price caps and cost-model benchmarking. At the same time, public utility legislation in America and Europe has been opening the telecommunications and electricity markets for competition but it has not been particularly deregulatory. On the contrary, the legislation imposes new obligations on dominant incumbents. The resulting regulation has been criticized as being asymmetric because it subjects incumbent or dominant carriers to more regulation than entrants and as incompatible with competition on equal terms (Schankerman, 1996). Asymmetry of regulation is really nothing new. It is an intrinsic part of regulation. In the past, the asymmetry was much more extreme in that it regulated entrants out of existence by erecting insurmountable legal entry barriers. What is more important now is to find the best strategy for substituting competition for regulation. We have argued above that deregulation is warranted in the horizontal competition context if actual competition in the market is sufficiently strong. This assessment is not easy, in particular if essential facilities persist.

Deregulation can occur in three broad stages.

First, regulation could be applied to all services of dominant incumbent carriers but price structures could be deregulated. This suggestion, first made by Vogelsang and Finsinger (1979) was later extended by L-T in the form of global price caps.

Second, there can be partial deregulation, leaving specific market segments to continuing regulation. Regulation could, in particular, retreat from competitive market segments (as done in the U.S. and Europe) or from retail markets altogether. Regulation would concentrate on bottleneck services and universal service issues. One of the nice features of price caps is that they can accommodate gradual deregulation. This can take two convenient forms. One is the reduction of regulated services by reducing the tariff basket.

The other is the permission to offer optional tariffs in addition to the price-cap regulated tariffs. Partial deregulation affects the remaining regulated areas, for example, because the price cap baskets may be reduced and the X-factor may require adaptation. Typically, the services to be deregulated have strongly declining prices so that the X-factor may have to be reduced. In case of deregulation of end-user services the regulation of essential inputs has to trace costs better. Thus, a move from usage-based to to capacity-based charging and non-linear pricing may be warranted.

Third and most extreme, there could be total deregulation. This is the route almost taken by New Zealand. It relies on antitrust policy rather than sector-specific regulation.[39] The problem with this approach is that it does not deal well with bottleneck issues in local telephone networks and with the highly political issue of universal service. That is why New Zealand has maintained price moratoria for local residential telephone services. In 1989, when the FCC introduced price caps for AT&T, they were viewed by many as an ideal bridge from regulation to unregulated competition. That proved to be correct, since AT&T's prices were deregulated in 1995. However, the same did not hold for price cap regulation of the LECs and all other cases of price caps.

Total deregulation appears to be unrealistic at this time for the main regulated industries, telecommunications and electricity, given the laws in place and the continuing dominance of incumbent carriers. The current choice is therefore between partial and global price caps. It may be surprising that both, extension of regulation to all services and contracting regulation to less services, can be deregulatory. The reason is in the form of regulatory constraints. If regulation fixes individual prices an extension of regulation to more market segments imposes stricter regulation. If, however, price cap regulation with freedom of price structure is applied, regulation of more market segments can be deregulatory. To see this, assume that a carrier offers two services, one in a monopoly market and one in a market where it faces (some) competition. If only the monopoly market is regulated under a price cap its maximum price is uniquely determined and all the carrier can do is stay below the cap. This is not much freedom if regulation is effective. In the second market, the carrier is constrained by the market price (which may contain profits, due to oligopolistic interaction). Now, assume that the regulator includes in the price cap the carrier's price in the second market and sets it at the current market price. Now, the carrier would gain some freedom because it could trade off a change in one price against a change in the other price in the opposite direction. But is this real freedom, given that the regulated price in the second market is constrained by the market price? First note that the firm is not made worse off because it could leave both prices where they were before. Second, however, the price in market 2 was profit maximizing to begin with. That means, the firm's profit function was flat at this point. In contrast, the price in market 1 was constrained by regulation. Hence, the firm's profit function was increasing in this price. Thus, by the envelope theorem, the firm can increase its profits by raising the price in market 1 and lowering the price in market 2.[40] So, the regulated firm would actually have gained additional freedom.

---

[39] See Vogelsang (1997).

[40] For changes in relative costs between the two markets global caps reduce the risks, because rebalancing is always possible. In contrast, if only market 1 is under price cap a relative cost increase in this market and

While this paper started out with a fairly general approach to incentive regulation of public utilities, it ended up concentrating on the role of price caps and yardstick regulation as regulatory tools to cope with problems of competition. A common view among economists seems to be that regulation interferes with competition and that this results in some trap. I have heard economists say, "Regulators don't want to deregulate for lack of competition, but competition does not materialize because of regulation." More recently, this phrase would be changed to, "Regulators don't want to deregulate because they believe that competition only exists because of regulation." The current paper contradicts the first of these phrases for the case of price cap regulation. It conjectures that the reduction in entry and market share of rivals caused by the aggressive behavior of an incumbent utility under price cap regulation could well result in a better market outcome than under unregulated competition. The same does not hold for the second phrase because global price cap regulation has not yet been applied to the combination of end-user rates and access charges. My conjecture is that it may be too late for global price caps because, at this stage, all players either prefer separate price caps or partial deregulation. Rivals prefer separate price caps, because the essential facilities basket is likely to take TELRIC/TSLRIC as the starting point. Incumbents prefer partial deregulation, because it is a first step towards full deregulation and because it may include stronger commitment power on the side of the regulator.

The current paper has made a prima facie case for some general rules. That does not mean that these general rules always lead to good results. If the danger from exceptions, such as strategic anticompetitive behavior, is viewed as severe, insurance policies may be needed. Such policies can take the form of imputation rules or price bands that are likely to restrict a regulated firm's behavior only in case of undesirable behavior.

---

a cost reduction in market 2 will lead to a (short or medium term) loss for the utility, because it has to lower its price in market 2 without the ability to raise the price in market 1. Conversely, if the costs in market 2 increase and decrease in market 1 the partially regulated firm will experience a windfall in market 1, while it can raise its price in market 2.

References

Abel, Jaison (1999), “Dominant Firm Pricing With Price-Cap Regulation and Fringe Competition: An Economic Analysis of Local Telephone Pricing”, NRRI/Ohio State University mimeo, December.

Acton, Jan, and Ingo Vogelsang (1989), "Introduction" to "Symposium on Price-Cap Regulation," *RAND Journal of Economics* 20, pp. 369-372.

Armstrong, Mark (2001), “The Theory of Access Pricing and Interconnection”, in Martin Cave, Sumit Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics.* Amsterdam: Elsevier Publishers, (forthcoming).

Armstrong, Mark, Simon Cowan, and John Vickers (1994), *Regulatory Reform: Economic Analysis and British Experience*. Cambridge, MA: MIT Press.

Armstrong, Mark, Ray Rees, and John Vickers (1995), “Optimal Regulatory Lag Under Price Cap Regulation”, *Revista Espanola De Economia* 10, pp. 93-116.

Averch, Harvey and Leland Johnson (1962), “Behavior of the Firm Under Regulatory Constraint,” *American Economic Review* 52, pp.1053-1069.

Baron, David P., and Raymond R. De Bondt (1979), “Fuel Adjustment Mechanisms and Economic Efficiency,” *Journal of Industrial Economics* 27, pp. 243-261.

Baron, David and Roger Myerson (1982), “Regulating a Monopolist with Unknown Costs”, *Econometrica* 50, pp. 911-930.

Baumol, William J. (1967), “Reasonable Rules for Rate Regulation: Plausible Polices for an Imperfect World,” in A. Phillips and O. E. Williamson (eds.), *Prices: Issues in Theory, Practice, and Public Policy*, Philadelphia: University of Pennsylvania Press, pp. 108-123.

Baumol, William J. (1982), “Productivity Adjustment Clauses and Rate Adjustment for Inflation,” *Public Utilities Fortnightly*, July 22, pp. 11-18.

Baumol, William J., W.F. Koehn and Robert D. Willig (1987), “How Arbitrary is Arbitrary, or: Towards the Desired Demise of Full Cost Allocation,” *Public Utilities Fortnightly* 120(5), September 3, pp.16-21.

Beesley, Michael and Stephen Littlechild (1989), “The Regulation of Privatized Monopolies in the United Kingdom,” *Rand Journal of Economics* 20, pp. 454-472.

Bernstein, Jeffrey and David Sappington (1999), “Setting the X Factor in Price Cap Regulation Plans,” *Journal of Regulatory Economics* 16, pp. 5-25.

Bradley, Ian (1993), “Price-Cap Regulation and Market Definition,” *Journal of Regulatory Economics* 3, pp.337-347.

Bradley, Ian and Catherine Price (1988), “The Economic Regulation of Private Industries by Price Constraints,” *Journal of Industrial Economics* 37, pp. 99-106.

Bradley, Ian and Catherine Price (1991), “Average Revenue Regulation and Regional Price Structure,” *Regional Science and Urban Economics* 21, pp. 89-108.

Braeutigam, Ronald R. (1980), “An Analysis of Fully Distributed Cost Pricing in Regulated Utilities,” *Bell Journal of Economics* 11, pp.182-196.

Brennan, Timothy (1989), “Regulating by ‘Capping Prices’,” *Journal of Regulatory Economics* 1, pp. 133-147.

Brennan, Timothy (1990), “Cross-Subsidization and Cost Misallocation by Regulated Monopolists,” *Journal of Regulatory Economics* 2, pp. 37-52.

Brown, Lorenzo, Michael Einhorn and Ingo Vogelsang (1989), “Incentive Regulation: A Research Report,” Federal Energy Regulatory Commission, Office of Economic Policy, Technical Report 89-3, November.

Brown, Lorenzo, Michael Einhorn and Ingo Vogelsang (1991), "Toward Improved and Practical Incentive Regulation," *Journal of Regulatory Economics* 3, pp. 323-338.

Cabral, Luis and Michael Riordan (1989), “Incentives for Cost Reduction Under Price Cap Regulation,” *Journal of Regulatory Economics* 1, pp. 93-102.

Cave, Martin (2000), “Price-Capping Mechanisms,” in Tim Jenkinson (ed.), *Readings in Microeconomics* (*Second Edition*). Oxford: Oxford University Press, pp. 224-238.

Cowan, Simon (1997), “Price-Cap Regulation and Inefficiency in Relative Pricing”, *Journal of Regulatory Economics* 12, pp. 53-70.

Crew, Michael and Paul Kleindorfer (1996), “Incentive Regulation in the United Kingdom and the United States: Some Lessons,” *Journal of Regulatory Economics* 9 pp. 211-225.

Crew, Michael, Paul Kleindorfer and Sudit (1979), Incentives for Efficiency in the Nationalised Industries: Beyond the 1978 White Paper,” *The Journal of Industrial Affairs* 7, pp. 11-15.

Diewert, W.E., and K.J. Fox (2000), “Incentive Indexes for Regulated Industries,” *Journal of Regulatory Economics* 17, pp. 5-24.

Donald, Stephen and David Sappington (1995), “Explaining the Choice Among Regulatory Plans in the U.S. Telecommunications Industry,” *Journal of Economics and Management Strategy* 4, pp. 237-265.

Donald, Stephen and David Sappington (1997), “Choosing Among Regulatory Options in the United States Telecommunications Industry,” *Journal of Regulatory Economics* 12, pp. 227-243.

Farrell, Joseph, and Bridger M. Mitchell (1998), “Benchmarking and the Effects of ILEC Mergers,” Declaration, October 14.

Farrell, Joseph, and Bridger M. Mitchell (1999), “”Response to Some Criticisms of Benchmarking Analysis,” April 9.

Faulhaber, Gerald R. (1975), “Cross-subsidization: Pricing in Public Enterprises,” *American Economic Review* 65, pp. 966-977.

Fraser, Rob (1995), “The Relationship Between the Costs and the Prices of a Multi-Product Monopoly: The Role of Price-Cap Regulation,” *Journal of Regulatory Economics* 8, pp. 23-32.

Fuss, Melvyn A., and Leonard Waverman (forthcoming), “Econometric Cost Functions in Telecommunications: Evidence on Scale, Scope, the Presence of Natural Monopoly, Technological Change and Efficiency: What Has It Told Us?” in Martin Cave, Sumit Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics.* Amsterdam: Elsevier Publishers, (forthcoming).

Goldberg, Victor P. (1976), “Regulation and Administered Contracts,” *Bell Journal of Economics* 7, pp.426-448.

Isaac, R. Mark (1991), “Price Cap Regulation: A Case Study of Some Pitfalls of Implementation,” *Journal of Regulatory Economics* 3, pp. 193-210.

Kridel, Donald, David Sappington, and Dennis Weisman (1996), "The Effects of Incentive Regulation in the Telecommunications Industry: A Survey", *Journal of Regulatory Economics* 9, pp. 269-306.

Laffont, Jean-Jacques (1994), "The New Economics of Regulation Ten Years After," *Econometrica*, 62, pp. 507-538.

Laffont, Jean-Jacques and Jean Tirole (1986), "Using Cost Observation to Regulate Firms", *Journal of Policitical Economy* 94, pp. 614-641.

Laffont, Jean-Jacques and Jean Tirole (1993), *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: The MIT Press.

Laffont, Jean-Jacques and Jean Tirole (1996), "Creating Competition Through Interconnection: Theory and Practice", *Journal of Regulatory Economics* 10, pp. 227-256.

Laffont, Jean-Jacques and Jean Tirole (2000), *Competition in Telecommunications*. Cambridge, MA: The MIT Press.

Lehman, Dale E., and Dennis L. Weisman (2000), "The Political Economy of Price Cap Regulation," *Review of Industrial Organization* 16(4), pp.343-356.

Levy, Brian, and Pablo Spiller (eds.) (1996), *Regulation, Institutions and Commitment: Comparative Studies of Telecommunications*, Cambridge: Cambridge University Press.

Liston, Catherine (1993), "Price-Cap versus Rate-of-Return Regulation," *Journal of Regulatory Economics* 5, pp. 25-48.

Littlechild, Stephen C. (1983), *Regulation of British Telecommunications' Profitability*, Report to the Secretary of State, Department of Industry, London: Her Majesty's Stationery Office.

Loeb, Martin, and Wesley A. Magat (1979), "A Decentralized Method of Utility Regulation," *Journal of Law and Economics* 22, pp.399-404.

Mitchell, Bridger M. (1990) "Incremental Costs of Telephone Access and Use", Report R-3909-ICTF, RAND Corporation, Santa Monica, July.

Neu, Werner (1993), "Allocative Inefficiency Properties of Price-Cap Regulation," *Journal of Regulatory Economics* 5, pp.159-182.

Newbery, David M. (2000), *Privatization, Restructuring, and Regulation of Network Industries*, Cambridge, MA, and London: MIT Press.

Sappington, David (1980), "Strategic Firm Behavior Under a Dynamic Regulatory Adjustment Process," *Bell Journal of Economics* 11, 1980, pp. 360-372.

Sappington, David (1983), "Optimal Regulation of a Multiproduct Monopoly with Unknown Technological Capabilities," *Bell Journal of Economics* 14, pp.453-463.

Sappington, David (2001), "Price Regulation and Incentives," in Martin Cave, Sumit Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics.* Amsterdam: Elsevier Publishers, (forthcoming).

Sappington, David and David Sibley (1992), "Strategic Nonlinear Pricing Under Price Cap Regulation," *Rand Journal of Economics* 23, pp. 1-19.

Schankerman, Mark (1996), "Symmetric Regulation for Competitive Telecommunications," *Information Economics and Policy* 8, pp.3-23.

Sharkey, William (2001), “Representation of Technology and Production,” in Martin Cave, Sumit Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics.* Amsterdam: Elsevier Publishers, (forthcoming).

Shin, Richard and John Ying (1993), “Efficiency in Regulatory Regimes: Evidence From Price Caps,” Presented at the Twenty-First Annual Telecommunications Policy Research Conference: Solomons, Maryland, September.

Shleifer, Andrei (1985), “A Theory of Yardstick Competition”, *Rand Journal of Economics* 16, pp. 319-327.

Spulber, Daniel F. (2001), “Competition Policy in Telecommunications,” in Martin Cave, Sumit Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics.* Amsterdam: Elsevier Publishers, (forthcoming).

Stigler, George, and Claire Friedland (1962), “What Can Regulators Regulate? The Case of Electricity,” *Journal of Law and Economics* 5, pp. 1-16.

Sudit, E. Fred (1979), “Automatic Rate Adjustments Based on Total Factor Productivity Performance in Public Utility Regulation,” in M. A. Crew (ed.), *Problems in Public Untility Economics and Regulation*, Lexington: Lexington Books, pp. 55-71.

Vogelsang, Ingo (1988), “A Little Paradox in the Design of Regulatory Mechanisms,” *International Economic Review* 29: 467-476.

Vogelsang, Ingo, (1989), "Price Cap Regulation of Telecommunications Services: A Long-Run Approach", in M.A. Crew (ed.), *Deregulation and Diversification of Utilities*, Boston: Kluwer Academic Publishers, pp. 21-42.

Vogelsang, Ingo (1997), “Antitrust vs. Sector Specific Approaches in Regulating Telecommunications Markets”, in D. Elixmann and P. Kuerble (eds.), *Multimedia - Potentials and Challenges from an Economic Perspective*, WIK Proceedings: Bad Honnef, pp.67-88.

Vogelsang, Ingo (Forthcoming), “Price Regulation for Independent Transmission Companies,” *Journal of Regulatory Economics*.

Vogelsang, Ingo, and J. Finsinger (1979), "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," *Bell Journal of Economics* 10, pp. 157-171.

Williamson, Oliver E. (1976), Franchise Bidding for Natural Monopolies – In General and with Respect to CATV,” *Bell Journal of Economics* 7, pp. 73-104.

Wimmer, Bradley S., und Gegory L. Rosston (2000), "From C to Shining C: Competition and Cross-subsidy in Communications," Paper presented at the 28th Telecommunications Policy Research Conference, September 24, Alexandria, Virginia.

Woroch, Glenn A. (2001), “Local Network Competition,” in Martin Cave, Sumit Majumdar, and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics.* Amsterdam: Elsevier Publishers, (forthcoming).